

Package ‘FBCanalysis’

March 26, 2022

Title Develop and evaluate time series data models based on fluctuation based clustering

Version 0.0.0.9000

Date 2020-03-16

Description The package includes tools for performing fluctuation-based clustering (FBC) on biological time series data, primarily for monitoring data fluctuations in asthmatics. The package includes functions for registering and processing time series data, developing matching clustering models based on Earth Mover's distances, and evaluating the models through enrichment analysis, stability after random data removal, or other frequently used cluster stability metrics.

License MIT + file LICENSE

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.1.2

URL <https://github.com/MrMaximumMax/FBCanalysis>

BugReports <https://github.com/MrMaximumMax/FBCanalysis/issues>

Depends R (>= 4.1.1)

Imports arsenal,
cluster,
clValid,
dplyr,
emdist,
FCPS,
glmnet,
imputeTS,
lubridate,
mclust,
RankAggreg,
readr,
stats,
tibble,
utils

Suggests rmarkdown,
knitr

VignetteBuilder knitr

R topics documented:

add_clust2enrich	2
add_clust2ts	3
add_enrich	4
clust_matrix	4
clValid_flow	5
emd_heatmap	6
emd_matrix	6
enr_obs_clust	7
init_clValid	8
jaccard_run_cognate	8
jaccard_run_emd	9
max_fluc	10
patient_boxplot	11
patient_hist	11
patient_list	12
patient_ts_plot	13
rnd_dat_rm	13
sim_jaccard_cognate	14
sim_jaccard_emd	15
sim_sample_enr	15
znorm	16
Index	17

add_clust2enrich	<i>Add clustering assignments to enrichment data frame</i>
------------------	--

Description

Add clustering assignments to enrichment data frame

Usage

```
add_clust2enrich(enrich, clustdat)
```

Arguments

enrich	Preprocessed enrichment data frame (also see function: add_enrich)
clustdat	Object of type list storing clustering data (also see function: clust_matrix)

Value

Processed enrichment data frame with added column indicating cluster assignments

Examples

```
list <- patient_list(
  "https://raw.githubusercontent.com/MrMaximumMax/FBCanalysis/master/demo/phys/data.csv",
  GitHub = TRUE)
#Sampling frequency is supposed to be daily
clustering <- clust_matrix(matrix, method = "kmeans", nclust = 3)
enr <- add_enrich(list,
  'https://raw.githubusercontent.com/MrMaximumMax/FBCanalysis/master/demo/enrich/enrichment.csv')
enr <- add_clust2enrich(enr, clustering)
```

add_clust2ts	<i>Add clustering assignments to time series data</i>
--------------	---

Description

Add clustering assignments to time series data

Usage

```
add_clust2ts(plist, clustdat)
```

Arguments

plist	List storing patient time series data (also see function: patient_list)
clustdat	Object of type list storing clustering data (also see function: clust_matrix)

Value

Processed data frame storing time series data with added column indicating cluster assignments

Examples

```
list <- patient_list(
  "https://raw.githubusercontent.com/MrMaximumMax/FBCanalysis/master/demo/phys/data.csv",
  GitHub = TRUE)
#Sampling frequency is supposed to be daily
clustering <- clust_matrix(matrix, method = "kmeans", nclust = 3)
ts <- add_clust2ts(list, clustering)
```

add_enrich	<i>Add enrichment data and preprocess for analysis</i>
------------	--

Description

Add enrichment data and preprocess for analysis

Usage

```
add_enrich(plist, path)
```

Arguments

plist	List storing patient time series data (also see function: patient_list)
path	Path where enrichment csv file is stored

Value

Processed data as object of type data frame; Enrichment data Patient_IDs are matched with Time Series Data List Patient IDs; In case it was indicated, NA values in the enrichment data are filled up by random sampling

Examples

```
list <- patient_list(
  "https://raw.githubusercontent.com/MrMaximumMax/FBCanalysis/master/demo/phys/data.csv",
  GitHub = TRUE)
#Sampling frequency is supposed to be daily
enr <- add_enrich(list,
  'https://raw.githubusercontent.com/MrMaximumMax/FBCanalysis/master/demo/enrich/enrichment.csv')
```

clust_matrix	<i>Cluster Earth Mover Distance Square Matrix data</i>
--------------	--

Description

Cluster Earth Mover Distance Square Matrix data

Usage

```
clust_matrix(matrix, method, nclust, plotclust)
```

Arguments

matrix	Object of type matrix storing Earth Mover's Distances for patient time series data distribution pairs
method	Clustering method (hierarchical, kmeans, diana, fanny, som, modelbased, sota, pam, clara)
nclust	Number of clusters (if not specified, user will be asked in the terminal)
plotclust	TRUE/FALSE if clustering data should be visualized (TRUE by default)

Value

Object of type list storing cluster data and clustering assignments for the Patient_IDs from the Earth Mover's Distance matrix

Examples

```
list <- patient_list(
  "https://raw.githubusercontent.com/MrMaximumMax/FBCanalysis/master/demo/phys/data.csv",
  GitHub = TRUE)
#Sampling frequency is supposed to be daily
matrix <- emd_matrix(list, "FEV1")
clustering <- clust_matrix(matrix, method = "hierarchical", nclust = 2)
```

clValid_flow	<i>Interactive console workflow to calculate and evaluate cluster validation measures</i>
--------------	---

Description

Interactive console workflow to calculate and evaluate cluster validation measures

Usage

```
clValid_flow(matrix, par)
```

Arguments

matrix	Earth Mover's Distance Matrix for processed patient time series data (also see functions: emd_matrix , patient_list)
par	Object of type list storing clustering methods and cluster range of interest; initialized via function: init_clValid

Value

Object of type list storing chosen clustering method and number of clusters (can be then used for function [clust_matrix](#))

Examples

```
list <- patient_list(
  "https://raw.githubusercontent.com/MrMaximumMax/FBCanalysis/master/demo/phys/data.csv",
  GitHub = TRUE)
#Sampling frequency is supposed to be daily
dismat <- emd_matrix(list, "PEF", maxIter = 5000)
parameters <- init_clValid()
output <- clValid_flow(dismat, parameters)
```

emd_heatmap	<i>Visualize an Earth Mover's Distance Square Matrix as a heatmap</i>
-------------	---

Description

Visualize an Earth Mover's Distance Square Matrix as a heatmap

Usage

```
emd_heatmap(input, parameter, maxIter)
```

Arguments

input	Earth Mover's Distance Matrix or list storing patient time series data (also see function: patient_list)
parameter	In case list is input, the parameter of interest from time series data list
Iter	In case input is time series data list, incate maxIter to calculate EMD matrix (also see function: emd_matrix)

Value

Visualized Earth Mover's Distance Matrix as a heatmap

Examples

```
list <- patient_list(
  "https://raw.githubusercontent.com/MrMaximumMax/FBCanalysis/master/demo/phys/data.csv",
  GitHub = TRUE)
#Sampling frequency is supposed to be daily
matrix <- emd_matrix(list, "FEV1")
emd_heatmap(matrix)
```

emd_matrix	<i>Generate a Earth Mover's Distance Matrix out of time series data list entries</i>
------------	--

Description

Generate a Earth Mover's Distance Matrix out of time series data list entries

Usage

```
emd_matrix(plist, parameter, maxIter)
```

Arguments

plist	List storing patient time series data (also see function: patient_list)
parameter	Parameter of interest to determine Earth Mover's Distances between distributions
maxIter	Maximum of iterations to calculate Earth Mover's Distance (default: 500)

Value

Earth Mover's Distance Square Matrix of type matrix

Examples

```
list <- patient_list(
  "https://raw.githubusercontent.com/MrMaximumMax/FBCanalysis/master/demo/phys/data.csv",
  GitHub = TRUE)
#Sampling frequency is supposed to be daily
matrix <- emd_matrix(list, "FEV1")
```

enr_obs_clust	<i>Observe specific cluster for overview and p-values</i>
---------------	---

Description

Observe specific cluster for overview and p-values

Usage

```
enr_obs_clust(ts.dat, enrich, clustno)
```

Arguments

ts.dat	Processed data frame storing time series data and cluster assignments (also see function: add_clust2ts)
enrich	Processed data frame storing enrichment data and cluster assignments (also see function: add_clust2enrich)
clustno	Cluster number of interest

Value

Terminal output presenting summary of time series and enrichment data with corresponding p-values

Examples

```
list <- patient_list(
  "https://raw.githubusercontent.com/MrMaximumMax/FBCanalysis/master/demo/phys/data.csv",
  GitHub = TRUE)
#Sampling frequency is supposed to be daily
clustering <- clust_matrix(matrix, method = "kmeans", nclust = 3)
enr <- add_enrich(list,
  'https://raw.githubusercontent.com/MrMaximumMax/FBCanalysis/master/demo/enrich/enrichment.csv')
enr <- add_clust2enrich(enr, clustering)
ts <- add_clust2ts(list, clustering)
enr_obs_clust(ts, enr, 1)
```

<code>init_clValid</code>	<i>Initialize Cluster Validation Measure Analysis in the context of Fluctuation Based Clustering (FBC) analysis</i>
---------------------------	---

Description

Initialize Cluster Validation Measure Analysis in the context of Fluctuation Based Clustering (FBC) analysis

Usage

```
init_clValid()
```

Value

Object of type list storing cluster method(s) and number of cluster range of interest (to be used for function: [clValid_flow](#))

Examples

```
init_clValid()
```

<code>jaccard_run_cognate</code>	<i>Simulate random data removal from time series data list and determine Jaccard index via Cognate Cluster approach for multiple random data removal steps</i>
----------------------------------	--

Description

Simulate random data removal from time series data list and determine Jaccard index via Cognate Cluster approach for multiple random data removal steps

Usage

```
jaccard_run_cognate(
  plist,
  parameter,
  n_simu,
  method,
  clust_num,
  n_clust,
  range
)
```


Arguments

plist	Object of type list storing patient time series data (also see function: patient_list)
parameter	Parameter of interest in time series data list
n_simu	Number of simulations
method	Clustering method (also see function: clust_matrix)
clust_num	Cluster of interest
n_clust	Number of clusters
range	Range to simulate random data removal (e.g. <code>c(0.1,0.2,0.5,0.7,0.8)</code>)

Value

Object of type list storing Jaccard indices for each indicated random data removal step and visualized results in a boxplot

Examples

```
list <- patient_list(
  "https://raw.githubusercontent.com/MrMaximumMax/FBCanalysis/master/demo/phys/data.csv",
  GitHub = TRUE)
output <- jaccard_run_cognate(list, "PEF", 10, "hierarchical", 1, 3, c(0.005, 0.01, 0.05, 0.1, 0.2))
```

jaccard_run_emd	<i>Simulate random data removal from time series data list and determine Jaccard index via Earth Mover's Distance approach for multiple random data removal steps</i>
-----------------	---

Description

Simulate random data removal from time series data list and determine Jaccard index via Earth Mover's Distance approach for multiple random data removal steps

Usage

```
jaccard_run_emd(plist, parameter, n_simu, method, clust_num, n_clust, range)
```

Arguments

plist	Object of type list storing patient time series data (also see function: patient_list)
parameter	Parameter of interest in time series data list
n_simu	Number of simulations
method	Clustering method (also see function: clust_matrix)
clust_num	Cluster of interest
n_clust	Number of clusters
range	Range to simulate random data removal (e.g. <code>c(0.1,0.2,0.5,0.7,0.8)</code>)

Value

Object of type list storing Jaccard indices for each indicated random data removal step and visualized results in a boxplot

Examples

```
list <- patient_list(
  "https://raw.githubusercontent.com/MrMaximumMax/FBCanalysis/master/demo/phys/data.csv",
  GitHub = TRUE)
#Sampling frequency is supposed to be daily
output <- jaccard_run_emd(list,"PEF",10,"hierarchical",1,3,c(0.005,0.01,0.05,0.1,0.2))
```

max_fluc	<i>Determine pair of maximum fluctuation difference in a list storing time series data</i>
----------	--

Description

Determine pair of maximum fluctuation difference in a list storing time series data

Usage

```
max_fluc(input, parameter, maxIter)
```

Arguments

input	Either a list storing time series data or EMD martrix (also see functions: patient_list , emd_matrix)
parameter	Parameter of interest from time series data list
maxIter	Maximum of iterations to apply for calculation of Earth Mover's Distannce (also see function: emd_matrix)

Value

Console output with Patient_ID pair, corresponding Earth Mover's Distance and visualized boxplot of both time series data distributions

Examples

```
list <- patient_list(
  "https://raw.githubusercontent.com/MrMaximumMax/FBCanalysis/master/demo/phys/data.csv",
  GitHub = TRUE)
#Sampling frequency is supposed to be daily
max_fluc(list, "PEF")
```

patient_boxplot	<i>Visualize patient(s) time series data in a boxplot for indicated parameter</i>
-----------------	---

Description

Visualize patient(s) time series data in a boxplot for indicated parameter

Usage

```
patient_boxplot(plist, patients, parameter, normalized)
```

Arguments

plist	List storing patient time series data (also see function: patient_list)
patients	Patient_ID(s) referring to (a) list element; can be single ID or multiple IDs (also see function: patient_list)
parameter	Parameter of interest in list element(s)
normalized	TRUE/FALSE if z-normalized (TRUE by default)

Value

Visualized patient(s) time series data in a boxplot for indicated parameter

Examples

```
list <- patient_list(
  "https://raw.githubusercontent.com/MrMaximumMax/FBCanalysis/master/demo/phys/data.csv",
  GitHub = TRUE)
#Sampling frequency is supposed to be daily
patient_boxplot(list,c("ID_2","testpat_1","testpat_2","a301"), "FEV1")
```

patient_hist	<i>Visualize patient time series data in a histogram for indicated parameter</i>
--------------	--

Description

Visualize patient time series data in a histogram for indicated parameter

Usage

```
patient_hist(plist, Patient_ID, parameter, normalized)
```

Arguments

plist	List storing patient time series data (also see function: patient_list)
Patient_ID	Patient_ID referring to a list element (also see function: patient_list)
parameter	Parameter of interest in list element
normalized	TRUE/FALSE if z-normalized (TRUE by default)

Value

Visualized patient time series data in a histogram for indicated parameter

Examples

```
list <- patient_list("
https://raw.githubusercontent.com/MrMaximumMax/FBCanalysis/master/demo/phys/data.csv",
GitHub = TRUE)
#Sampling frequency is supposed to be daily
patient_hist(list,"testpat_1","PEF")
```

patient_list	<i>Process patient time series data by interpolation options and store data in an object of type list</i>
--------------	---

Description

Process patient time series data by interpolation options and store data in an object of type list

Usage

```
patient_list(path, GitHub)
```

Arguments

path	Path where csv file(s) are stored (only folder, not specific file(s))
GitHub	Set TRUE when csv file comes from GitHub (FALSE by default)

Value

Object of type list storing patient time series data

Examples

```
list <- patient_list(
"https://raw.githubusercontent.com/MrMaximumMax/FBCanalysis/master/demo/phys/data.csv",
GitHub = TRUE)
#Sampling frequency is supposed to be daily
```

patient_ts_plot	<i>Visualize patient time series data in a time series plot for indicated parameter</i>
-----------------	---

Description

Visualize patient time series data in a time series plot for indicated parameter

Usage

```
patient_ts_plot(plist, Patient_ID, parameter, normalized)
```

Arguments

plist	List storing patient time series data (also see function: patient_list)
Patient_ID	Patient_ID referring to a list element (also see function: patient_list)
parameter	Parameter of interest in list element
normalized	TRUE/FALSE if z-normalized (TRUE by default)

Value

Visualized patient time series data in a time series plot for indicated parameter

Examples

```
list <- patient_list(
  "https://raw.githubusercontent.com/MrMaximumMax/FBCanalysis/master/demo/phys/data.csv",
  GitHub = TRUE)
#Sampling frequency is supposed to be daily
patient_ts_plot(list,"testpat_1","PEF")
```

rnd_dat_rm	<i>Remove random data from time series data list</i>
------------	--

Description

Remove random data from time series data list

Usage

```
rnd_dat_rm(plist, removal)
```

Arguments

plist	Object of type list storing patient time series data (also see function: patient_list)
removal	Amount of data removal (0 = 0%, 1 = 100%)

Value

Object of type list storing patient time series data with indicated amount of data removed randomly

Examples

```
list <- patient_list(
  "https://raw.githubusercontent.com/MrMaximumMax/FBCanalysis/master/demo/phys/data.csv",
  GitHub = TRUE)
#Sampling frequency is supposed to be daily
list_rm <- rnd_dat_rm(testlist, 0.95)
```

sim_jaccard_cognate	<i>Simulate random data removal from time series data list and determine Jaccard index via Cognate Cluster approach</i>
---------------------	---

Description

Simulate random data removal from time series data list and determine Jaccard index via Cognate Cluster approach

Usage

```
sim_jaccard_cognate(plist, parameter, removal, n_simu, method, n_clust, Iter)
```

Arguments

plist	Object of type list storing patient time series data (also see function: patient_list)
parameter	Parameter of interest in time series data list
removal	Amount of random data removal to determine Jaccard index
n_simu	Number of simulations
method	Clustering method (also see function: clust_matrix)
n_clust	Number of clusters (also see function: clust_matrix)
Iter	Maximum iterations to determine Earth Mover's Distances (also see function: emd_matrix); default is 5,000 for this function

Value

Object of type matrix storing received Jaccard indices for indicated amount of random data removal for all clusters

Examples

```
list <- patient_list(
  "https://raw.githubusercontent.com/MrMaximumMax/FBCanalysis/master/demo/phys/data.csv",
  GitHub = TRUE)
#Sampling frequency is supposed to be daily
output <- sim_jaccard_cognate(list, "PEF", 0.05, 10, "hierarchical", 2, 1000)
```

sim_jaccard_emd	<i>Simulate random data removal from time series data list and determine Jaccard index via Earth Mover's Distance approach</i>
-----------------	--

Description

Simulate random data removal from time series data list and determine Jaccard index via Earth Mover's Distance approach

Usage

```
sim_jaccard_emd(plist, parameter, removal, n_simu, method, n_clust, Iter)
```

Arguments

plist	Object of type list storing patient time series data (also see function: patient_list)
parameter	Parameter of interest in time series data list
removal	Amount of random data removal to determine Jaccard index
n_simu	Number of simulations
method	Clustering method (also see function: clust_matrix)
n_clust	Number of clusters (also see function: clust_matrix)
Iter	Maximum iterations to determine Earth Mover's Distances (also see function: emd_matrix); default is 5,000 for this function

Value

Object of type matrix storing received Jaccard indices for indicated amount of random data removal for all clusters

Examples

```
list <- patient_list(
  "https://raw.githubusercontent.com/MrMaximumMax/FBCanalysis/master/demo/phys/data.csv",
  GitHub = TRUE)
output <- sim_jaccard_emd(list, "PEF", 0.05, 10, "hierarchical", 2, 100)
```

sim_sample_enr	<i>Simulate random sampling for NA entries in enrichment data and check stability of resulting p-values for the enrichment parameters</i>
----------------	---

Description

Simulate random sampling for NA entries in enrichment data and check stability of resulting p-values for the enrichment parameters

Usage

```
sim_sample_enr(plist, path, clustdat, clustno, n_sim)
```

Arguments

plist	List storing patient time series data (also see function: patient_list)
path	Path where enrichment csv file is stored
clustdat	Object of type list storing clustering data (also see function: clust_matrix)
clustno	Cluster number of interest
n_sim	Number of simulations

Value

Object of type list storing the received p-values for each parameter in a vector and boxplot visualizing the received p-values

Examples

```
list <- patient_list(
  "https://raw.githubusercontent.com/MrMaximumMax/FBCanalysis/master/demo/phys/data.csv",
  GitHub = TRUE)
#Sampling frequency is supposed to be daily
path <- 'https://raw.githubusercontent.com/MrMaximumMax/FBCanalysis/master/demo/enrich/enrichment.csv'
test <- sim_sample_enr(list,path,clustering,1,100)
sim_sample_enr <- function(plist, path, clustdat, clustno, n_sim)
```

znorm

*z-normalise data***Description**

Function applicable on any numeric distribution of data and z-normalize them.

Usage

```
znorm(data)
```

Arguments

data	Numeric distribution; may be stored in an object of type vector, matrix or data frame
------	---

Value

z-normalized distribution of input

Examples

```
random_distribution <- runif(n = 50, min = 1, max = 10)
znorm_distribution <- znorm(random_distribution)
```


Index

add_clust2enrich, [2](#), [7](#)
add_clust2ts, [3](#), [7](#)
add_enrich, [2](#), [4](#)

clust_matrix, [2](#), [3](#), [4](#), [5](#), [9](#), [14–16](#)
clValid_flow, [5](#), [8](#)

emd_heatmap, [6](#)
emd_matrix, [5](#), [6](#), [6](#), [10](#), [14](#), [15](#)
enr_obs_clust, [7](#)

init_clValid, [5](#), [8](#)

jaccard_run_cognate, [8](#)
jaccard_run_emd, [9](#)

max_fluc, [10](#)

patient_boxplot, [11](#)
patient_hist, [11](#)
patient_list, [3–6](#), [9–11](#), [12](#), [13–16](#)
patient_ts_plot, [13](#)

rnd_dat_rm, [13](#)

sim_jaccard_cognate, [14](#)
sim_jaccard_emd, [15](#)
sim_sample_enr, [15](#)

znorm, [16](#)